Supplementary Material for Orderless Recurrent Models for Multi-label Classification

Vacit Oguz Yazici[1,2], Abel Gonzalez-Garcia[1], Arnau Ramisa[2,3],
Bartlomiej Twardowski[1], Joost van de Weijer[1]
[1] Computer Vision Center, Universitat Autonoma de Barcelona, Barcelona, Spain
[2] Wide-Eyes Technologies, Barcelona, Spain [3] Universitat de Vic, Barcelona, Spain
{voyazici,agonzalez,bartlomiej.twardowski,joost}@cvc.uab.es, arnau.ramisa@uvic.cat

1. Introduction

This is the supplementary material of the paper *Orderless Recurrent Models for Multi-label Classification*. We first exhibit some images whose labels are sorted differently by the proposed *predicted label alignment* (PLA) and frequent-first approach. Then, we show co-occurrence matrices of different super-categories of the MS-COCO dataset computed by the best BCE and LSTM models.

2. Qualitative comparison of PLA and other methods

In Figures 1-2, we can see different orders yielded by the PLA and frequent-first approaches. The images are chosen to emphasize the problems with the approaches that use predefined orders. As can be seen in the images, the frequent-first approach always predicts the labels in the same order. This leads to confusion in case of dominant but less frequent objects or minor but more frequent objects in an image. Then, this confusion leads to duplicate predictions in different time steps.

3. Additional co-occurrence matrices of LSTM and BCE models

In Figure 3, the predicted class co-occurence matrices for binary cross entropy (BCE) and predicted label alignment (PLA) models can be seen. The levels of co-occurence in BCE are noticeably higher than those on PLA, as it reuses the same parts of image for different predictions of similar objects (e.g. bike and motorbike). This can be observed especially on the animals, food, vehicle and kitchen super-categories. In the animals super-category the BCE model overshoots co-occurrence of dogs-cats and horses-cows, while in the vehicles the confusion is on the buses, trucks and cars. In the kitchen super-category the confusion is the worst since most of the images are images of entire kitchens and the BCE model uses the entire scene for different predictions. On the other hand, the LSTM model has much lower differences with the ground truth, since the previous predictions are taken into account at every time step.

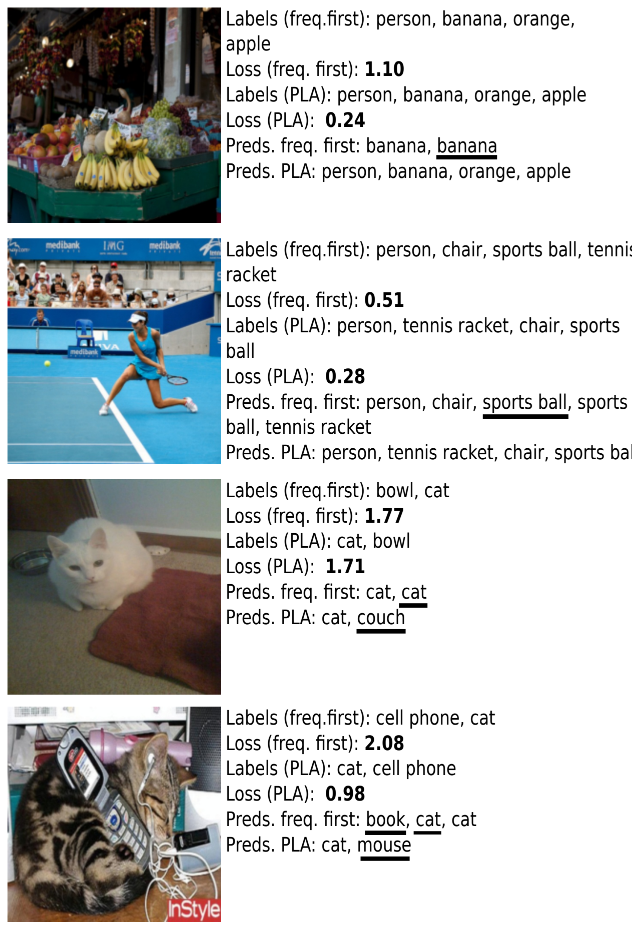
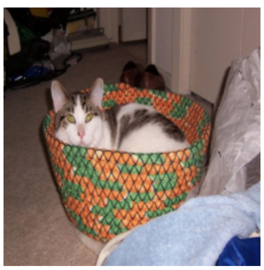


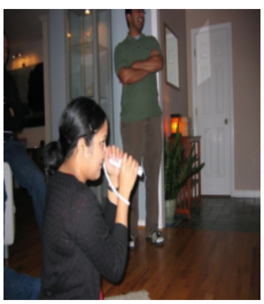
Figure 1: Comparisons of orders yielded by the PLA and frequent-first approaches (wrong or duplicate predictions are underlined).

Labels (freq.first): person, skis
Loss (freq. first): **0.54**
Labels (PLA): person, skis
Loss (PLA): **0.23**
Preds. freq. first: person, skis, skis
Preds. PLA: person, skis

Labels (freq.first): person, umbrella
Loss (freq. first): **3.35**
Labels (PLA): person, umbrella
Loss (PLA): **2.13**
Preds. freq. first: person, chair, bench
Preds. PLA: person, chair, umbrella



Labels (freq.first): cat
Loss (freq. first): **8.66**
Labels (PLA): cat
Loss (PLA): **1.58**
Preds. freq. first: person, cat
Preds. PLA: cat, bed



Labels (freq.first): person, couch, potted plant, remote
Loss (freq. first): **1.07**
Labels (PLA): person, remote, potted plant, couch
Loss (PLA): **0.78**
Preds. freq. first: person, couch, remote
Preds. PLA: person, remote, potted plant, couch

Labels (freq.first): person, chair, sports ball, tennis racket
Loss (freq. first): **0.36**
Labels (PLA): person, tennis racket, sports ball, chair
Loss (PLA): **0.16**
Preds. freq. first: person, chair, tennis racket, tennis racket

Labels (freq.first): teddy bear
Loss (freq. first): **1.02**
Labels (PLA): teddy bear
Loss (PLA): **0.09**
Preds. freq. first: bear
Preds. PLA: teddy bear

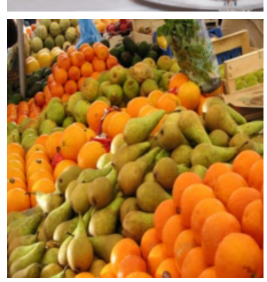
Labels (freq.first): dining table, pizza
Loss (freq. first): **0.89**
Labels (PLA): pizza, dining table
Loss (PLA): **0.26**
Preds. freq. first: pizza
Preds. PLA: pizza, dining table

Labels (freq.first): person, surfboard, kite
Loss (freq. first): **0.82**
Labels (PLA): person, kite, surfboard
Loss (PLA): **0.37**
Preds. freq. first: person, kite
Preds. PLA: person, kite, surfboard

Labels (freq.first): person, sports ball, tennis racket
Loss (freq. first): **0.47**
Labels (PLA): person, tennis racket, sports ball
Loss (PLA): **0.10**
Preds. freq. first: person, tennis racket
Preds. PLA: person, tennis racket, sports

Labels (freq.first): person, car, frisbee
Loss (freq. first): **0.85**
Labels (PLA): person, frisbee, car
Loss (PLA): **0.23**
Preds. freq. first: person, frisbee, frisbee
Preds. PLA: person, frisbee, car

Labels (freq.first): person, dining table, knife, tie, cake
Loss (freq. first): **0.94**
Labels (PLA): person, tie, cake, dining table, knife
Loss (PLA): **0.32**
Preds. freq. first: person, knife, knife, cake
Preds. PLA: person, tie, cake, dining table, knife



Labels (freq.first): orange, apple
Loss (freq. first): **0.57**
Labels (PLA): orange, apple
Loss (PLA): **0.37**
Preds. freq. first: orange
Preds. PLA: orange, apple

Figure 2: See caption of Figure 1.

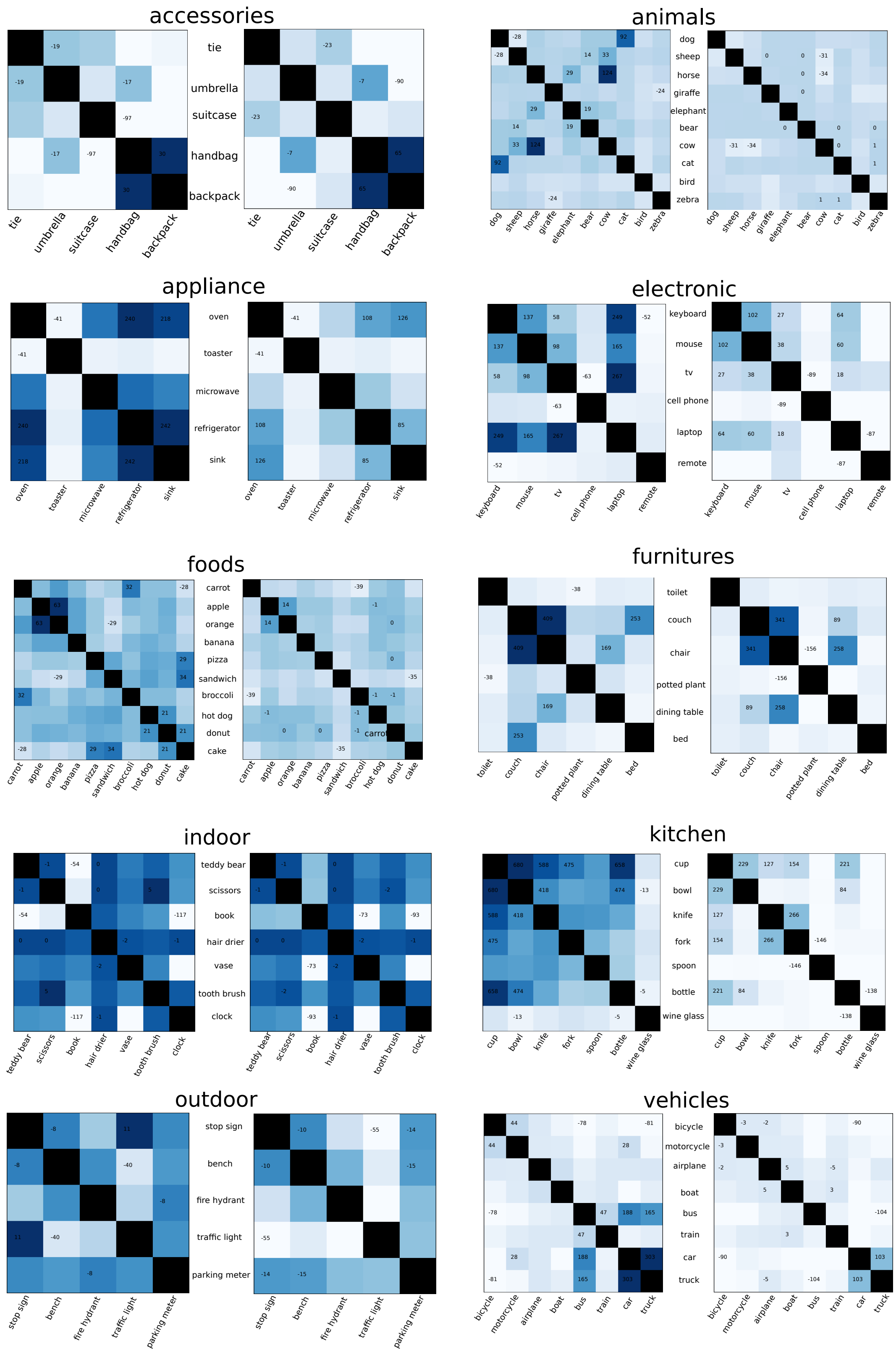


Figure 3: Co-occurence matrices for BCE (left) and PLA (right) models. BCE re-uses evidence to predict different objects, and hence has higher co-occurence levels due to false positives.